¹ The role of cross-linguistic lexical similarity on bilingual word acquisition

² Gonzalo Garcia-Castro[1], Daniela Avila-Varela[1], & Nuria Sebastian-Galles[1]

³ [1] Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain

⁴ Author Note

⁵ Campus de Ciutadella, Universitat Pompeu Fabra, 08005, Barcelona, Spain

⁶ Correspondence concerning this article should be addressed to Gonzalo

⁷ Garcia-Castro, Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain. E-mail:

⁸ gonzalo.garciadecastro@upf.edu

Abstract

Bilinguals face the challenging task of learning words from languages with overlapping phonologies. Floccia et al. (2018) reported larger vocabulary sizes for 24-month-old bilinguals that were learning languages that shared a greater amount of cognates (e.g., English-Dutch). The mechanisms underlying this effect remain unknown. We explore two compatible scenarios. First, we test whether cognates are learnt earlier than non-cognates. This would account for the difference in vocabulary size associated to the amount of shared cognates across languages. Second, we explore the possibility that the word-forms of one language interact with those form the other language, scaffolding the acquisition of their translation equivalents when their phonologies overlap. This mechanism, in line with the parallel activation account of bilingual speech perception, would provide a plausible explanation to why cognates are acquired ealier by bilinguals. We developed an online tool to collect parental reports of receptive and productive vocabularies from children learning Catalan and/or Spanish, and present data on receptive and productive vocabulary of bilingual toddlers aged 12 to 34 months.

Keywords: lexical acquisition, vocabulary, bilingualism

Word count: X

²⁶ The role of cross-linguistic lexical similarity on bilingual word acquisition

²⁷ **Introduction**

²⁸ **Introduction**

²⁹ Before the end of their first year of life, infants start directing their gaze to some
³⁰ object when hearing their labels, according to both experimental data (Bergelson &
³¹ Swingley, 2012; Jusczyk & Aslin, 1995; Tincoff & Jusczyk, 1999) and parental reports (e.g.,
³² Fenson et al., 1994). During the last half of their second year, they acquire new words at
³³ an increasingly fast rate (Bergelson, 2020; Bloom, 2002; Fenson et al., 1994). Despite the
³⁴ considerable variation in children's trajectories of vocabulary growth, such trajectories
³⁵ seem quite consistent across languages (Braginsky, Yurovsky, Marchman, & Frank, 2019).
³⁶ Nonetheless, most of the literature on early word acquisition has been conducted on
³⁷ monolingual children, and does not address the problem of how bilinguals–who represent a
³⁸ substantial proportion of the population in most societies–acquire words at early ages.

³⁹ There is evidence that bilinguals know less words in each of their languages than
⁴⁰ monolinguals, but also that both groups know a similar amount of words when the two
⁴¹ languages are aggregated. For example, Hoff et al. (2012) found that English-Spanish
⁴² bilingual toddlers in South Florida knew less words in English than monolinguals (who
⁴³ only learnt English). In contrast, both groups knew a similar total amount of words, when
⁴⁴ both English and Spanish vocabularies were counted together. Other studies have provided
⁴⁵ converging evidence that bilinguals know a similar–or even larger–number of words than
⁴⁶ monolinguals, only when the languages are aggregated (Fabian, 2016; Gonzalez-Barrero,
⁴⁷ Schott, & Byers-Heinlein, 2020; Oller & Eilers, 2002; J. L. Patterson, 2004; J. Patterson,
⁴⁸ Pearson, & Goldstein, 2004; Pearson & Fernández, 1994; Pearson, Fernández, & Oller,
⁴⁹ 1993; Petitto et al., 2001; Smithson, Paradis, & Nicoladis, 2014). While these studies have
⁵⁰ mostly relied on samples of bilingual children learning two relatively distant languages (as

it is the case of English and Spanish) it is unclear whether children learning typologically more similar languages also know less words in each of their languages than monolinguals. What role could linguistic distance play during early vocabulary growth?

For a given set of concepts, bilingual children may be exposed to two distinct sets of word-forms–one in each language. Depending on the linguistic distance between both languages, the two sets of words may overlap in varying degrees. Particularly, when both languages are linguistically close, like Spanish and Catalan (both Roman languages), they are more likely to share a larger amount of cognates (i.e., form-similar translation equivalents) than two linguistically distant languages, like Spanish and English–one Roman, the other Germanic. For instance, in the presence of a door, a Spanish-Italian (or a Spanish-Catalan) bilingual might hear *puerta* and *porta* (cognates), whereas a Spanish-English bilingual might hear *puerta* and *door* (non-cognates). It could be the case that mapping two phonologically similar labels (cognates, like *puerta-porta*) onto the same referent is easier than doing the same with two phonologically dissimilar labels (non-cognates, like *puerta* and *door*). If cognates are easier to acquire than non-cognates, bilinguals learning a pair of languages that share a high proportion of cognates should benefit more often from this facilitation effect than those learning a pair of languages with a lower proportion of cognates, and should therefore show larger vocabulary sizes.

A recent study by Floccia et al. (2018) provided evidence in line with this claim. The authors collected vocabulary data on word comprehension and production from 372 24-month-old bilingual toddlers in the United Kingdom who were learning English and an additional language. The additional language was one a pool of 13 typologically diverse languages: Bengali, Cantonese Chinese, Dutch, French, German, Greek, Hindi/Urdu, Italian, Mandarin Chinese, Polish, Portuguese, Spanish and Welsh. The authors calculated the average phonological similarity between the words in each of these additional languages and their translation equivalents in English. Phonological similarity was measuring by computing the Levenshtein distance between each cross-language pair of phonological

78 transcriptions. The Levenshtein distance is a metric that computes the edit distance

79 between two strings by counting the smallest number of insertions, deletions and

80 substitutions one of the strings has to go through to become identical to the other

81 (Levenshtein & others, 1966). Floccia et al. divided the resulting score by the length of the

82 longest string to bound the similarity scores between 0 and 1, and then entered this variable

83 as a predictor as they modelled participants' vocabulary sizes. Among other findings, the

84 authors reported an increase in productive vocabulary size in the additional language

85 associated with an increase in the average phonological similarity between the translation

86 equivalents of each language pair. For example, English-Dutch bilinguals (22.14%

87 phonological similarity), were able to produce more Dutch words than English-Mandarin

88 bilinguals (1.97% phonological similarity) were able to produce in Mandarin.

89   Floccia et al. pointed to *parallel activation* as the main mechanism underpinning

90 their results. The parallel activation hypothesis suggests that bilinguals activate both

91 languages simultaneously during speech production or comprehension. This phenomenon is

92 the result of the activation of lexical representations in both languages, even when only one

93 is in use during production (Costa, Caramazza, & Sebastian-Galles, 2000; Hoshino & Kroll,

94 2008) or comprehension (Thierry & Wu, 2007). One of the clearest examples of parallel

95 activation was provided by Costa, Caramazza, and Sebastian-Galles (2000). In this study,

96 Catalan-Spanish monolingual and bilingual adults were asked to name pictures of common

97 objects in Spanish. In half of the trials, the object labels were cognates in Spanish and

98 Catalan (*árbol-arbre*, translations of *tree*), whereas in the other half of the trials labels were

99 non-cognates (*mesa-taula*, translations of *table*). Bilinguals named cognate pictures faster

100 than non-cognate pictures, even after adjusting for the lexical frequency of the items.

101 Importantly, Spanish monolinguals did not show this effect. These results suggest that, for

102 the bilinguals, Catalan phonology was activated during the production of Spanish words,

103 facilitating the naming of cognate pictures. Several recent studies have also provided

104 similar evidence in comprehension in children (Poulin-Dubois, Bialystok, Blaye, Polonia, &

Yott, 2013; e.g., Von Holzen & Mani, 2012). Parallel activation is therefore a plausible mechanism to account for Floccia et al.'s results: cognates increase the amount of activation in both languages, facilitating word acquisition. Therefore, children learning language pairs with a larger proportion of cognates are predicted to show larger vocabulary sizes. If cognates are the reason for this facilitation effect, it could be predicted that cognates are acquired earlier than non-cognates. Likely as it might seem, an earlier age of acquisition for cognates is not the only scenario where Floccia et al.'s results are account for. For example, languages that share more cognates might also share a larger overlap in their phonemic inventories, which could facilitate the acquisition of all words, regardless of their cognate status.

However, the evidence on an earlier age of acquisition for cognates compared to non-cognates is scarce. To the best of our knowledge only two studies have provided direct data on this issue. First, Bosch and Ramon-Casas (2014) used vocabulary parental reports (152 lexical items) from 48 Catalan-Spanish bilinguals aged 18 months, and found that cognates represented a larger proportion of participant's vocabulary than non-cognates. Second, Schelletter (2002) reported a longitudinal single case of one English-German bilingual who produced cognates earlier than non-cognates, on average. The low sample size in these studies make it challenging to draw strong conclusions about the effect of cognateness in vocabulary growth. Floccia et al.'s estimates are statically more reliable given their (much larger) sample size, but their study was not aimed at testing the effect of cognateness on age of acquisition directly. In their discussion the authors state the following (pp. 70):

"This finding also provides support to the proposal that the cognate advantage is due to cognates being acquired before non-cognates in early childhood (Costa et al., 2016), leading to an ease of processing later in life."

We identify two main reasons why we cannot conclude an earlier age of acquisition

for cognates than for non cognate from Floccia et al.'s results. First, the response variable used was the proportion of words each participant understood and/or produced, from the list of lexical items in the vocabulary checklists. By aggregating the responses from all items into a single datum per child, information about the acquisition status of cognates vs. non-cognates was lost. Second, all participants were aged 24 months, meaning that even if the unaggregated responses to individual items were included as response variable, the possible effect of cognateness could only be interpreted as an increase or decrease in the likelihood of participants at this age to have acquired each item, and not as an increase or decrease in the age of acquisition of such item. In summary, the evidence supporting an earlier age of acquisition of cognates vs. non-cognates presents some limitations that prevents drawing sound conclusions about this effect.

Another worth-mentioning finding in Floccia et al. is that the increase in vocabulary size associated with linguistic similarity was larger in the additional language vocabulary than in English vocabulary. Most participants in their sample were English-dominant, meaning that their relative amount of exposure to English was larger than in the additional language. Given that the relative amount of exposure to a given language is a strong predictor of children's vocabulary size in that language (e.g., Cattani et al., 2014; Thordardottir, Rothenberg, Rivard, & Naves, 2006), participants may have, on average, learnt the English word-form of translation equivalents earlier than the word-form in the additional language. If this is the case, then the acquisition of English words by English-dominant participants would rarely benefit from their cognate status (the other word-form is not available yet), while the acquisition of words in the additional language would benefit from their phonological similarity with the (available) English form.

The aim of this study is to investigate the role of cognateness in bilingual word acquisition. We hypothesised that cognate words would be acquired earlier than non-cognate word, and that this difference would be larger in participants with lower exposure to the language such word belongs to. Using an designed online vocabulary

checklist–carefully designed specifically for this study–we collected data from a sample of

bilinguals aged 10 to 36 months learning Catalan and/or Spanish, with varying degrees of

exposure to each language. We then modelled the probability of participants being

reported by their parents to understand and/or say each word in the checklist, conditional

to its cognate status in Catalan and Spanish and participants' degree of exposure to the

corresponding language, while adjusting for participants' age and the item's lexical

frequency.

References

Method

Results

Discussion

References

Bergelson, E. (2020). The comprehension boost in early word learning: Older

infants are better learners. *Child Development Perspectives*, *14*(3), 142–149.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the

meanings of many common nouns. *Proceedings of the National Academy of

Sciences*, *109*(9), 3253–3258.

Bloom, P. (2002). *How children learn the meanings of words.* MIT press.

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual

toddlers' expressive vocabulary: Does form similarity matter? *International

Journal of Behavioral Development*, *38*(4), 317–322.

https://doi.org/10.1177/0165025414532559

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, *38*(4), 317–322. https://doi.org/10.1177/0165025414532559

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67.

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to english is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language & Communication Disorders*, *49*(6), 649–671.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The Cognate Facilitation Effect: Implications for Models of Lexical Access. *Journal of Experimental Psychology*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The Cognate Facilitation Effect: Implications for Models of Lexical Access. *Journal of Experimental Psychology*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

Fabian, A. P. (2016). Investigating vocabulary abilities in bilingual portuguese-english-speaking children.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185.

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., … others. (2018). I: INTRODUCTION. *Monographs of the Society for Research in Child Development*, *83*(1), 7–29.

Gonzalez-Barrero, A. M., Schott, E., & Byers-Heinlein, K. (2020). Bilingual adjusted vocabulary: A developmentally-informed bilingual vocabulary measure.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, *106*(1), 501–511.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23.

Levenshtein, V. I., & others. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710). Soviet Union.

Oller, D. K., & Eilers, R. E. (2002). *Language and literacy in bilingual children* (Vol. 2). Multilingual Matters.

Patterson, J. L. (2004). Comparing bilingual and monolingual toddlers' expressive vocabulary size.

Patterson, J., Pearson, B., & Goldstein, B. (2004). Bilingual language development and disorders in spanish–english speakers.

Pearson, B. Z., & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*, *44*(4), 617–653.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in

bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43*(1), 93–120.

Petitto, L. A., Katerelos, M., Levy, B. G., Gauna, K., Tétreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language*, *28*(2), 453.

Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *International Journal of Bilingualism*, *17*(1), 57–70.

Schelletter, C. (2002). The effect of form similarity on bilingual children's lexical development. *Bilingualism: Language and Cognition*, *5*(2), 93–107. https://doi.org/10.1017/S1366728902000214

Smithson, L., Paradis, J., & Nicoladis, E. (2014). Bilingualism and receptive vocabulary achievement: Could sociocultural context make a difference. *Bilingualism: Language and Cognition*, *17*(4), 810–821.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530–12535.

Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, *4*(1), 1–21.

Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, *4*(1), 1–21.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in

6-month-olds. *Psychological Science*, *10*(2), 172–175.

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586.

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586.